Microbial Pathogenesis 1988; **4**: 345–357

Nucleotide sequence of the invasion plasmid antigen B and C genes (*ipa*B and *ipa*C) of *Shigella flexneri*

Bernadette Baudry, Michel Kaczorek* and Philippe J. Sansonetti†

Service des Entérobactéries, Unité INSERM 199, Institut Pasteur, 75724 Paris Cedex 15, France

(Received January 29, 1988; accepted February 15, 1988)

Baudry B. (Service des Entérobactéries, Unité INSERM 199, Institut Pasteur, 75724 Paris Cedex 15, France), M. Kaczorek and P. J. Sansonetti. Nucleotide sequence of the invasion plasmid antigen B and C genes (*ipa*B and *ipa*C) of *Shigella flexneri. Microbial Pathogenesis* 1988; **4**: 345–357. The nucleotide sequence of a 4.8 kilobase (kb) *Hind*III fragment from pWR100, the virulence plasmid of *Shigella flexneri* 5, was determined and analysed. This fragment encodes polypeptides b (62 kilodalton, kD) and c (43 kD) which have already been described as two of the four immunogenic polypeptides of *Shigellae*. The nucleotide sequence revealed that in addition to the *ipa*B and *ipa*C genes encoding polypeptides b and c, a third complete open reading frame was found within the fragment. The gene, named *ipp*I, encoded a 17 kD polypeptide. The deduced amino acids sequence of polypeptides b and c showed no signal peptide but presence of highly hydrophobic domains compatible with a transmembraneous location. The surprising A and T richness of the three genes as compared with the *Escherichia coli* and *Shigella* genomes, resulted in a biased codon usage, and raises the question of the origin of the sequences.

Key words: DNA sequence; *S. flexneri*; invasion genes.

Introduction

The pathogenic potential of *Shigellae*, the etiologic agents of bacillary dysentery, is correlated with the ability of these bacteria to enter and multiply within colonic epithelial cells.[1] It has now been well established that expression of genes located both on the virulence plasmid and on the chromosome is required for full virulence.[2–6] However, the plasmid by itself is sufficient to promote the entry of the bacteria into cells.[6]

Up to now, very few data have been published on proteins which could be involved in the invasive process. Nonetheless, seven plasmid-encoded polypeptides have been found to be specifically associated with invasive strains of *Shigella flexneri*.[7] Among these, four polypeptides, named a, b, c and d, were consistently recognized by sera from monkeys which had been infected with *S. flexneri* and represent the major proteinaceous antigens of *Shigellae*.[7] Another study has shown that sera from children recovering from shigellosis contained antibodies directed against the same

* Present address: Laboratoire de recherche et de dévelopement, Pasteur Vaccins, 92430 Marnes la Coquette, France.
† Corresponding author.

0882-4010/88/050345+13 $03.00/0

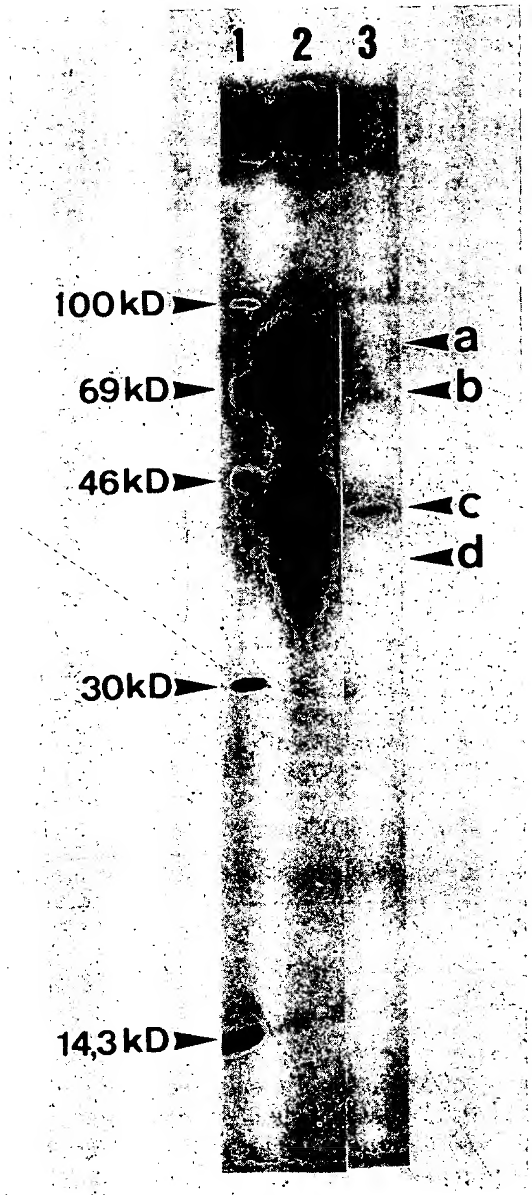


Fig. 1. Immunoblot of whole cell extracts reacted with serum from a monkey immunized against *Shigella flexneri* 2a (diluted 1/200). Lane 1, molecular weights (BRL); 2, M90T; 3, BS169/pHS5753. Exposure 6 days.

polypeptides.[8] Polypeptides b and c, with estimated molecular weights of 62 and 43 kilodaltons (kD) respectively, appear to be the dominant antigens.

We have recently reported the cloning[9] and the characterization by mutagenesis and subcloning[10] of a 45 kilobase (kb) fragment from pWR100, the virulence plasmid of *S. flexneri* serotype 5 strain M90T. This recombinant plasmid confers both the ability to invade HeLa cells, and the capacity to express the four immunogenic polypeptides.[9] Study of insertion mutants revealed that low expression of the immunogenic polypeptides b, c and d resulted in a dramatic decrease in the invasive ability of the mutant

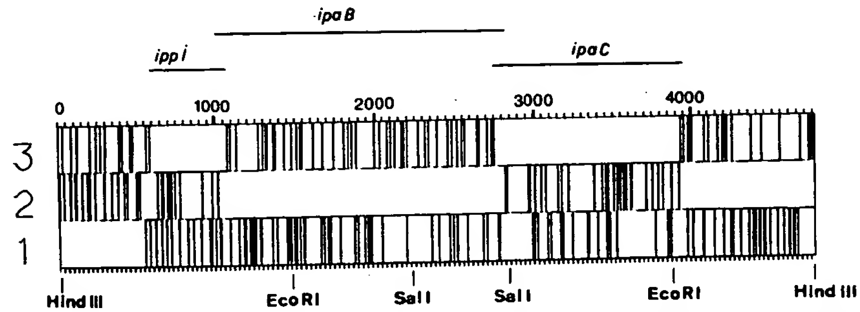


Fig. 2. Open reading frames (ORF) and restriction map of the 4.8 kb *Hind*III fragment. Vertical lines represent the nonsense codons on the total length of the 4.8 kb segment. ORF corresponding to genes *ipp*I, *ipa*B and *ipa*C are indicated. Scale is in base pairs.

strain, whereas a mutant which did not express polypeptide a was still invasive in the HeLa cell assay.[10] We have decided to focus on the study of polypeptides b and c because of their high immunogenicity which may have important implications for vaccine design.

This manuscript reports the complete amino acid sequence of polypeptides b and c deduced from the nucleotide sequence of genes *ipa*B and *ipa*C. We also report the presence of an open reading frame (ORF) located upstream of *ipa*B, which encodes a 17 kD polypeptide. The genetic organization of the genes is discussed.

Results

Expression of immunogenic polypeptides

A 4.8 kb *Hind*III fragment spanning the region encoding polypeptides b and c as predicted by insertion mutagenesis,[10] was subcloned into the *Hind*III site of plasmid pACYC184.[11] The resulting plasmid, pHS5753, was subsequently introduced into *Shigella* strain BS169.

To analyse the ability of the *Hind*III fragment to direct synthesis of polypeptides b and c, whole cell extracts from BS169 carrying plasmid pHS5753 were analysed by immunoblotting using antiserum from a monkey orally immunized against *S. flexneri* 2a. Figure 1 shows that both polypeptides were detected, migrating with a relative molecular weight (M_r) of 62 and 43 kD respectively. Therefore, it is likely that the 4.8 kb *Hind*III fragment from the virulence plasmid pWR100 of *S. flexneri* contains the *ipa*B and *ipa*C genes.

Nucleotide sequence

The nucleotide sequence of the 4.8 kb *Hind*III fragment was completely determined on both strands. Surprisingly, in addition to the two expected open reading frames (ORF) for polypeptides b and c, a third ORF was found (Fig. 2). All three ORFs were on the same DNA strand, and therefore in the same orientation. The ORFs were 485 bases pairs (bp) ("i"), 1802 bp ("b") and 1108 bp ("c") long. The complete nucleotide sequence of these three ORFs is represented in Fig. 3.

Furthermore, the nucleotide sequence showed the presence, on the same strand, of two other truncated ORFs located at each end of the cloned fragment (Fig. 2). At the 5′ end region, is a 547 bp long ORF which probably starts upstream of the *Hind*III cloning site. At the 3′ region of the 4.8 kb fragment, an ORF 842 bp long is interrupted

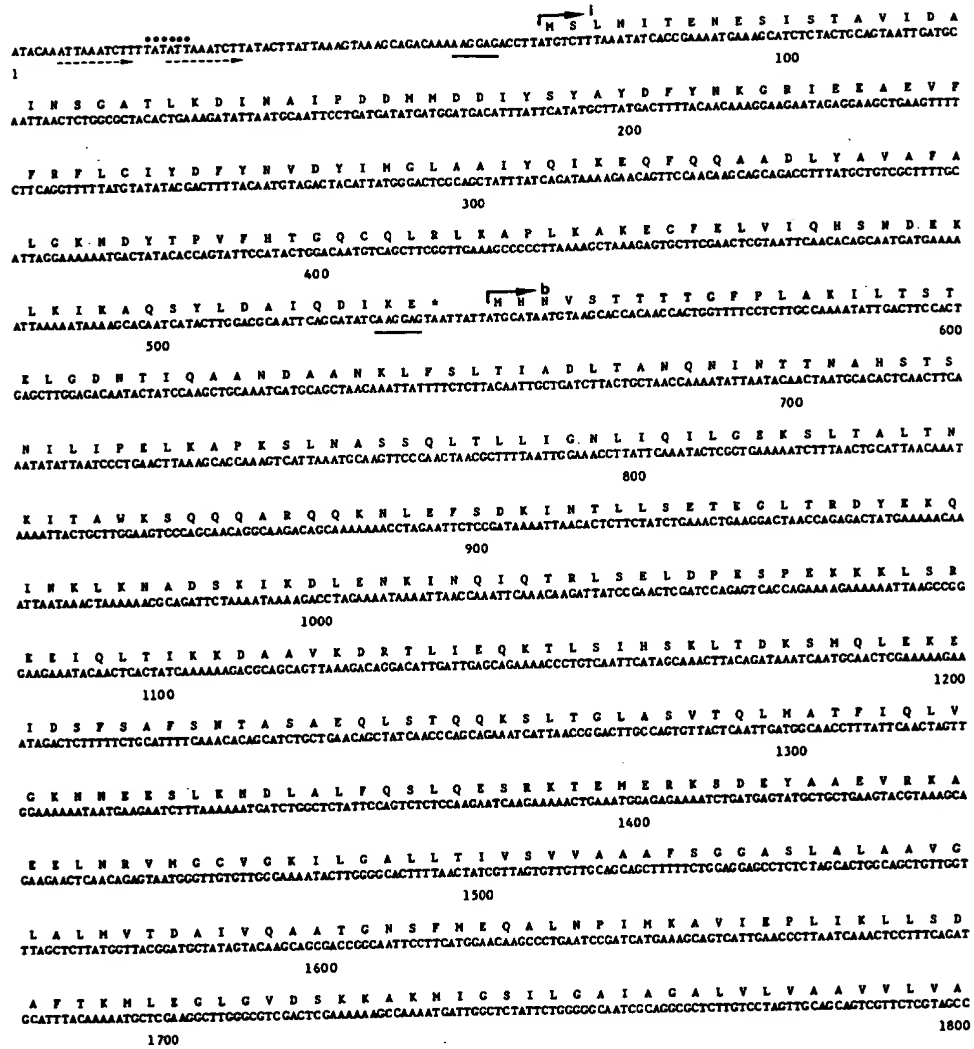


Fig. 3. Nucleotide sequence of the *ippI*, *ipaB* and *ipaC* genes and deduced peptide sequence of corresponding polypeptides i, b, and c. Arrows correspond to inverted repeats found in the nucleotide sequence, whereas broken arrows represent direct repeats. Possible RBS are underlined, and putative "−10" and "−35" sequences are signaled with dots.

by the other *Hind*III cloning site. Nucleotide sequences of these two truncated ORFs are not shown.

In addition to the *Sal* I site previously mapped,[10] a search for restriction endonuclease sites within the sequence revealed the presence of a second *Sal*I site which had not been detected by agarose gel electrophoresis when the physical map of pHS4108 was determined. The *Sal*I-*Sal* I fragment generated was 383 bp long and was easily detected by electrophoresis on polyacrylamide gel (data not shown).

Features of the nucleotide sequence

The G+C percentage, calculated either on the total DNA sequence, or on each complete ORF, was 37%.

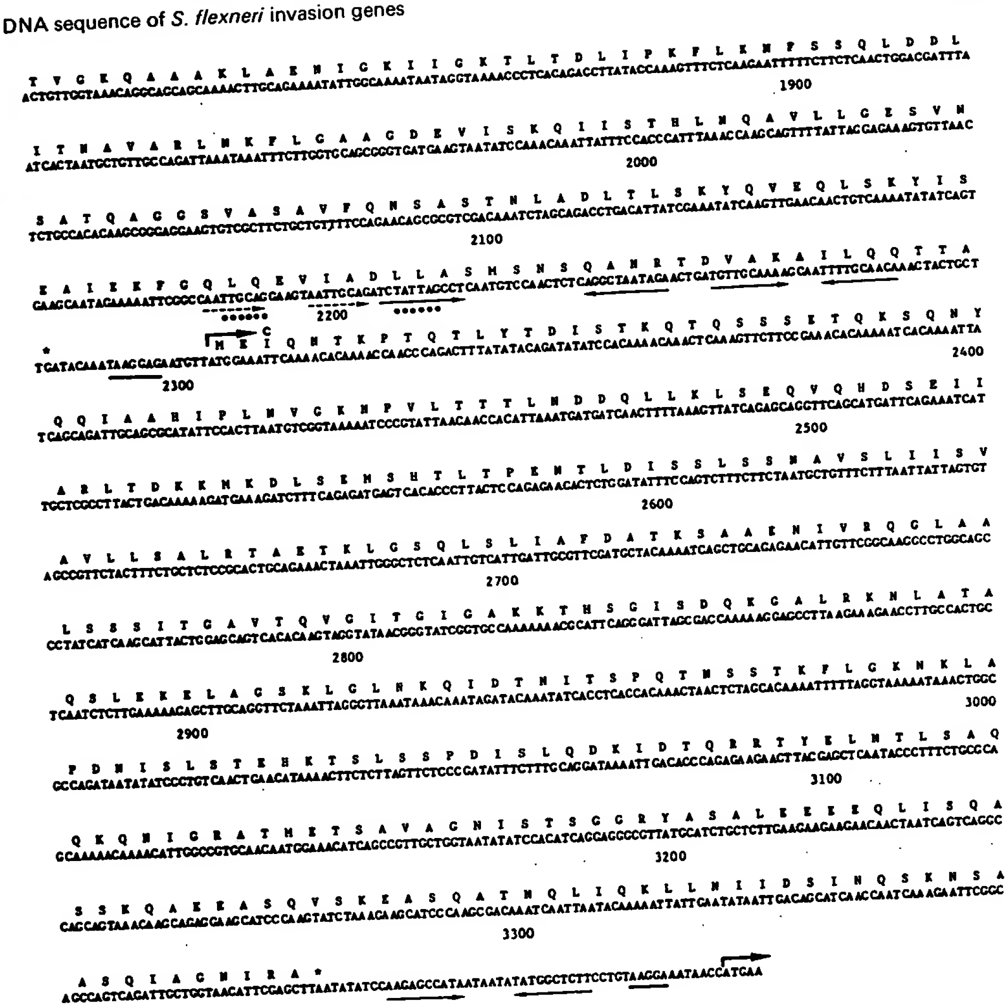


Fig. 3.—*continued*

Each ORF presented an ATG codon close to the beginning of the frame and was terminated either by a TAA or by a TGA stop codon. The first ATG was found at 21 bp and 63 bp respectively, downstream of the 5′ end of the ORFs "i" and "b". ORF "c" presented an ATG codon 70 bp from the 5′ end of the ORF.

The first ATG of ORF "i" and "b", and the second ATG of ORF "c", were closely preceded (position −11, −13 and −14 from the ATG respectively) by a Shine–Dalgarno sequence (6 or 7 bp long) complementary to the 3′ end of the 16S rRNA of *Escherichia coli*.[12] These short sequences represented good putative ribosome binding sites (RBS). Therefore, the first ATG codons found downstream were considered as the initiation codons for translation. Hence, proteins encoded by these three ORFs should have respective molecular weights of 17 736 (i), 62 130 (b), and 38 733 (c) daltons as calculated from the amino acid sequence deduced from the nucleotide sequence. The 3′ end truncated ORF was also preceded by a RBS sequence.

The M_r of polypeptides b and c had previously been estimated at 62 kD and 43 kD respectively on polyacrylamide gels.[7] Upon polypeptide size criteria and genetic and physical mapping data,[10] the *ipa*B gene was associated to ORF "b" and *ipa*C to the

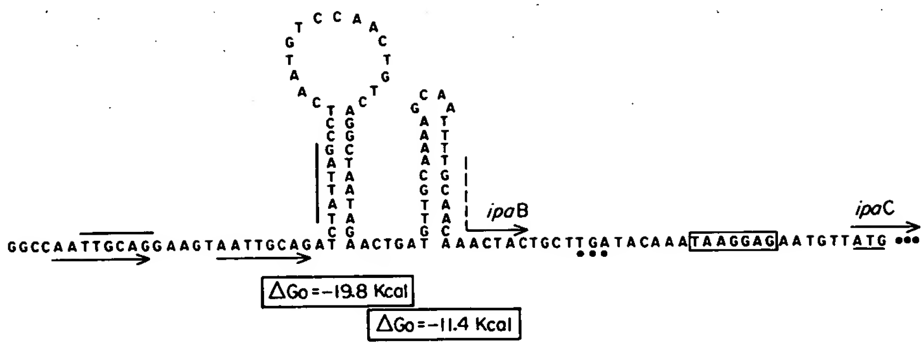


Fig. 4. Detail of the sequence showing possible loop structures at the 3′ end of the *ipa*B gene. Direct repeats are underlined with arrows. TGA termination codon of the *ipa*B gene is underlined with dots. The RBS sequence (box) and the initiation codon (underlined) of *ipa*C gene are represented. Thick lines show the position of the putative "−10" and "−35" sequences upstream of *ipa*C gene. Δ Go at 25°C of the loop structures is indicated and was calculated according to Cantor and Schimmel.[24]

ORF "c". The ORF "i" was associated with the newly identified polypeptide i (see below: "*identification of the ippI product*") and has been named *ipp*I (for invasion plasmid polypeptide).

According to taxonomic studies, *Shigellae* are very closely related to *E. coli*.[13] Moreover, cloned *Shigella* genes are usually easily expressed in *E. coli*.[14] Therefore, similarities between transcriptional and translational signals from *Shigella* and *E. coli* were expected. Upstream of the *ipa*C gene possible "−10" and "−35" regions were found. At position −96 from the ATG, a TATTAG sequence which is related to the "−10" *E. coli* consensus sequence TATAAT,[15] and at position −114 from the ATG, a TTGCAG sequence similar to the "−35" consensus sequence TTGACA of *E. coli*[15] were found (Fig. 3). The distance of 16 bp between these putative "−10" and "−35" sequences is comparable to the 17 bp found in the *E. coli* promoter regions. Upstream of the *ipp*I gene, a putative "−10" sequence, TATATT, was found at position −45 from the predicted initiation codon ATG, but no sequence related to the "−35" consensus sequence could be found. No sequences related either to the "−10" or "−35" promoter consensus sequences could be found within the 5′ non-coding region of *ipa*B.

Direct and inverted repeats, some of which are represented on Fig. 3, were found on the sequence. Figure 4 shows part of the sequence corresponding to the 3′ end of *ipa*B, which also contains the putative promoter sequence for the *ipa*C gene. Within this 100 bp segment a direct repeat, which comprises the possible "−35" of *ipa*C, and two inverted repeats, which could form stable loops, were found (Fig. 4). The putative "−10" sequence of the *ipa*C gene was located within one of these inverted repeats. Presence of such potential secondary structure within a promoter region has been previously reported for some genes regulated at the transcriptional level.[16]

On the complementary strand, another ORF was detected, which was 591 bp long. However, the first ATG was located in the middle of the ORF, 253 bp downstream. This ATG was also preceded by a putative RBS which was similar to the sequence previously seen in front of the other genes.

Identification of the ippI *gene product*

Polyacrylamide gel electrophoresis of [^{35}S]-methionine labeled proteins from minicells containing pHS5753 was performed to detect the product of the *ipp*I gene. As

expected, a polypeptide of ca. 17 kD was present in the extract from minicells containing pHS5753 (Fig. 5). The evaluated M_r of this polypeptide was in good agreement with the calculated molecular weight (M_w) of the product encoded by the *ipp*I ORF. Therefore, the 17 kD polypeptide was associated with the *ipp*I gene and named polypeptide i.

Codon usage

The mean frequency of utilization of codons in the three genes *ipp*I, *ipa*B and *ipa*C compared to the codon usage in *E. coli*[17] is presented in Table 1. As expected from

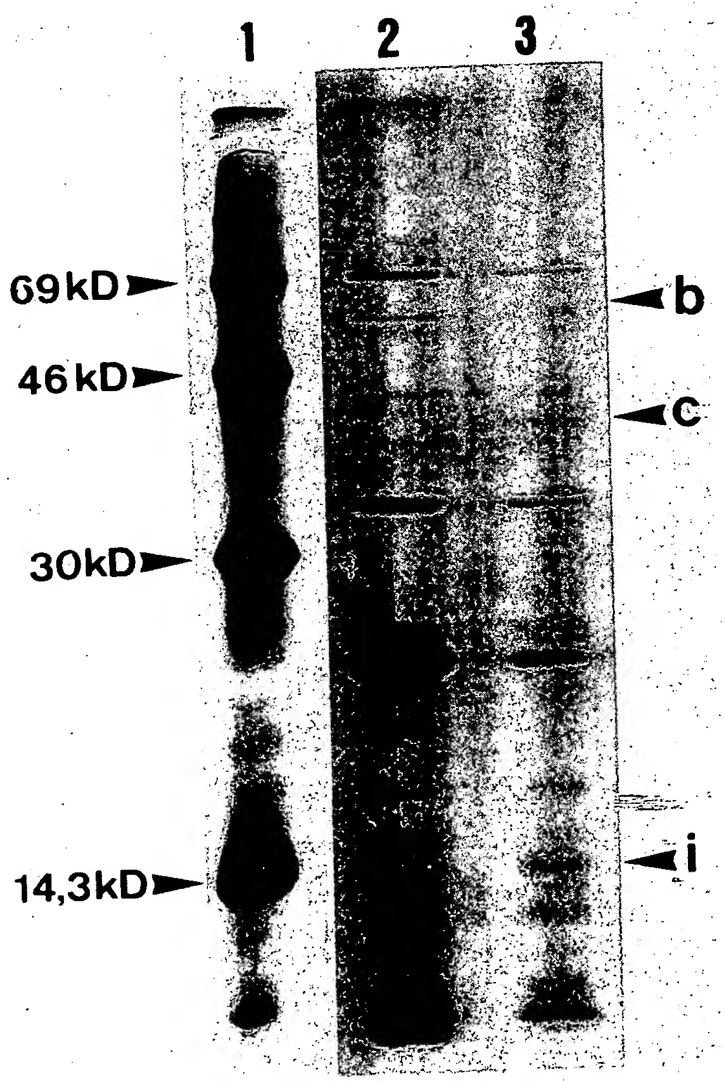


Fig. 5. Electrophoresis of [^{35}S] methionine labelled protein from minicells. Lane 1, molecular weights; 2, pACYC184; 3, pHS5753. Arrows indicate position of polypeptides b, c, and i.

Table 1 Comparison between the codon usage of *E. coli* genes and the ORF found within the *Hind*III fragment. Codon usage of *E. coli* is from[17]. Bold and underlined numbers respectively emphasize particularly high or low frequencies of utilization by the *ipp*I, *ipa*B and *ipa*C *Shigella* genes as compared with *E. coli* genes.

Amino acid	Codon	% *S. flexneri*	% *E. coli*	Amino acid	Codon	% *S. flexneri*	% *E. coli*
Arg	CGU	13.0	58.1	Val	GUU	51.1	37.5
	C	8.7	35.0		C	17.0	12.9
	A	4.3	2.3		A	31.9	22.9
	G	13.0	3.2		G	0.0	26.8
	AGA	**52.2**	1.2				
	G	8.7	0.3	Ile	AUU	52.9	37.3
					C	19.5	62.2
Leu	CUU	**26.0**	8.6		A	**27.6**	0.5
	C	11.8	6.6				
	A	12.6	1.8	Lys	AAA	88.6	76.7
	G	10.2	69.1		G	11.4	23.3
	UUA	**29.1**	5.8				
	G	10.2	8.2	Asn	AAU	**64.6**	24.2
					C	35.4	75.8
Ser	UCU	32.7	26.5				
	C	14.5	25.6	Gln	CAA	**66.7**	26.6
	A	**26.4**	8.3		G	33.3	73.4
	G	3.6	6.5				
	AGU	12.7	6.5	His	CAU	**72.7**	38.9
	C	10.0	21.6		C	27.3	61.1
Thr	ACU	412	23.8	Glu	GAA	77.3	73.4
	C	20.0	50.6		G	22.7	26.6
	A	**34.1**	5.9				
	G	4.7	19.7	Asp	GAU	60.4	51.0
					C	39.6	49.0
Pro	CCU	13.0	9.0				
	C	22.2	6.0	Tyr	UAU	68.4	40.6
	A	**55.6**	19.9		C	31.6	59.4
	G	5.6	65.1				
				Cys	UGU	**75.0**	42.0
Ala	GCU	29.8	27.9		C	25.0	58.0
	C	14.9	18.8				
	A	47.9	22.9	Phe	UUU	57.7	43.5
	G	7.4	30.5		C	42.3	56.5
Gly	GGU	32.2	47.8	Met	AUG	100.0	100.0
	C	22.0	40.8				
	A	**34.0**	4.6	Trp	UGG	100.0	100.0
	G	12.0	6.8				

the high A+T content of the genes, codon usage is quite different from that of *E. coli*. The frequencies show that codons ending in A or U are largely preferred (Table 1). A striking example is the AGA codon for arginine which is used at a frequency of 52.2% in these *Shigella* genes, and only 1.2% in *E. coli.*

Features of the polypeptides

The amino acid composition deduced from the sequence was rather similar for the three polypeptides. Briefly, polypeptide i contained 1.94% cysteine and all amino acids except tryptophan, polypeptide b contained all residues but only one cysteine (0.17%), and polypeptide c contained neither tryptophan nor cysteine.

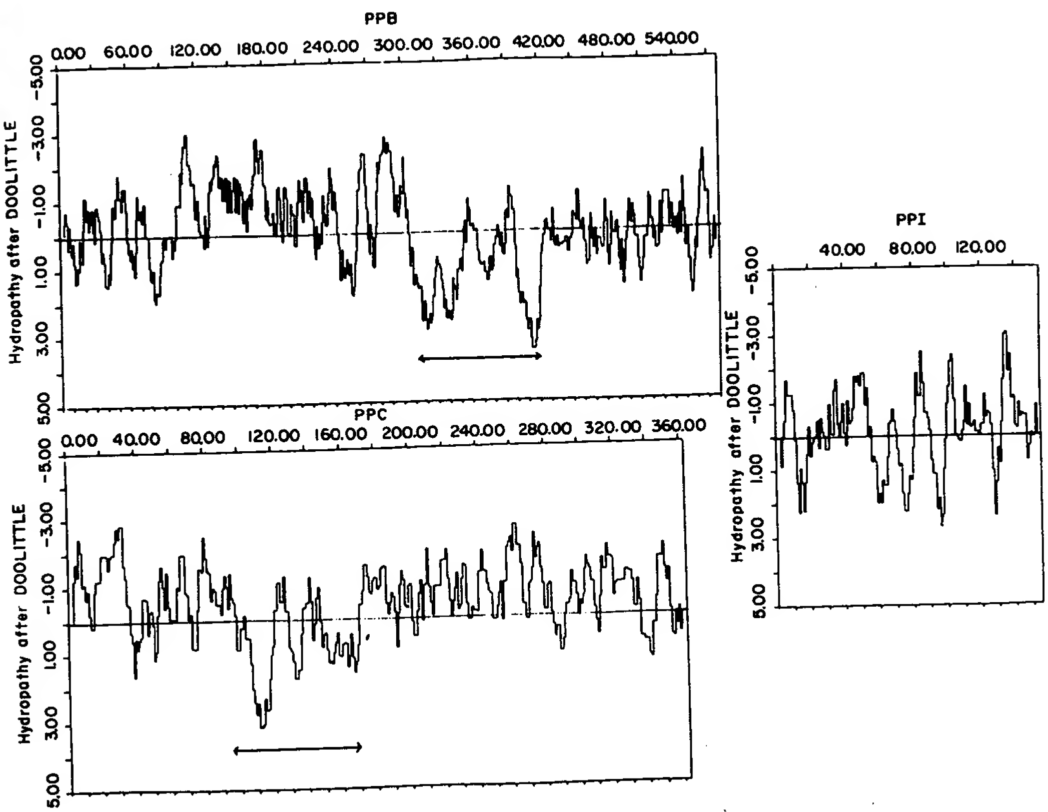


Fig. 6. Hydropathy profile of the deduced amino acid sequences of polypeptides i, b, and c, using the Kyte and Doolittle program[32] with windows of 6, 8 and 7 amino acids respectively. Hydrophobic regions of polypeptides b and c are indicated by an arrow.

The hydropathy profile of each polypeptide was calculated by using the HYDROPLOT program of Kyte and Doolittle[32] and is shown on Fig. 6. The hydropathy profile of polypeptide i showed no particular features. On the other hand, polypeptides b and c both appeared to contain a hydrophobic domain within the central part of the molecule. In polypeptide b, the hydrophobic domain, approximately 120 amino acids long, was closely preceded by a very hydrophilic region of 180 amino acids. In the case of polypeptide c, the ca. 60 amino acids long hydrophobic domain was immediately followed by an approximately 110 residues long region which was hydrophilic.

Discussion

In this study, we have determined the nucleotide sequence of a 4.8 kb *Hind*III fragment from the recombinant plasmid pHS4108.[9,10] An immunoblot confirmed that this fragment contained the *ipa*B and *ipa*C genes, encoding immunogenic polypeptides b and c respectively, as predicted by the analysis of insertion mutants.[10]

The 4.8 kb sequence consisted of five ORFs, with very short spaces in between. Three of these reading frames were entirely contained within the fragment whereas two others were truncated by the ends of the cloned segment. A 6 or 7 bp sequence complementary to the 3′ end of *E. coli* 16S rRNA was found close to the 5′ end of each

ORF. After this, the first following ATG codons were designated as the initiation codons.

Based upon ORF size criteria and previously obtained data on the position of the genes encoding the immunogenic polypeptides, the 1802 bp long ORF "b", and the 1178 bp long ORF "c" were attributed to genes *ipa*B and *ipa*C respectively. ORF "i" could encode a 17 kD polypeptide which had never been identified. Study of the products of pHS5753 indeed confirmed that a ca. 17 kD polypeptide was expressed. This polypeptide has been named i, and the corresponding ORF *ipp*I. However, no insertion mutants have ever been found within this gene, therefore, we do not know if it is involved in the invasion process.

Though promoter related sequences could not be very easily predicted, possible "−10" and "−35" sequences were found upstream of ORF "c", and interpreted as a putative promoter for the *ipa*C gene. No promoter-related sequences could be found in front of the gene *ipa*B. Though this fact does not necessarily mean that no promoter exists in front of *ipa*B, the possibility of a co-transcription of *ipp*I and *ipa*B could be considered and would raise a singular point. Indeed, it was observed on the sequence that the TAA codon marking the end of translation of polypeptide i was just preceded by the probable RBS for the translation of polypeptide b. The very short space between the two coding sequences (22 bp) and the absence of potential promoter-related sequences suggest that *ipp*I and *ipa*B may be transcribed on the same operon. This overlapping of translational (termination and initiation) signals has already been observed on the sequence of the operons encoding *Vibrio cholerae* enterotoxin,[18,19] and *E. coli* Shiga-like toxin.[20,21] In both cases, the overlapping features are thought to act as some sort of translation regulatory structure, responsible for the 1 to 5 ratio observed between the products encoded by the two genes.

Until now, polypeptides b and c have been considered to be present in the membrane, or in the periplasmic space, because enhanced signals could be seen on immunoblots of membrane preparations and because antiserum directed against b and c could be obtained by injecting rabbits with water extracts of *Shigella*. However, the hydropathy profiles of these polypeptides revealed no signal peptide structures which could be expected for secreted proteins.[22] On the other hand, both polypeptides contained a large internal hydrophobic segment that could be an intra-membraneous region.

Interesting results were obtained by comparing codon usage of the three genes and with that of *E. coli* genes. The A+T richness of the sequenced fragment resulted in a bias in codon usage: codons with an end in A or U were largely favoured. Up to now, only one gene of *Shigella* has been sequenced.[23] This gene, *vir*F, was cloned from the 220 kb virulence plasmid of *S. flexneri* 2a. The nucleotide sequence of *vir*F has revealed the same A+T richness and a preponderant use of codons ending in A or U as in the genes of the *Hind*III fragment. Although up to now the total length of sequenced segments represents only 7 kb out of the 220 kb virulence plasmid of *Shigellae*, it is very likely that these differences are conserved among the other virulence genes of the plasmids. These results raise interesting questions about the origin of the virulence genes carried by the plasmid of *Shigellae*. Furthermore, the difference of composition of the DNA may possibly explain the general instability of the virulence plasmid, and the high instability of the sequences responsible for invasion. The biased codon usage, resulting in utilization of rare tRNAs, might also explain why the polypeptides involved in invasion are so weakly produced.

Determination of the nucleotide sequence of the *ipa*B and *ipa*C genes opens new fields of research on the molecular mechanism of the virulence of *Shigellae*. Moreover, based on the amino acid sequence, studies of the immunogenic domains and antigenic epitopes of polypeptides b and c will be undertaken, and will be helpful in the search for a vaccine against shigellosis.

Table 2 Bacterial strains and plasmids

Strain or plasmid	Characteristics	Plasmid content	Reference
S. flexneri 5 M90T	wild type	pWR100 2 small cryptic plasmids	5
S. flexneri 2a			
BS169	Mal$^+$ λ^s *galU*::Tn*10*	2 small cryptic plasmids	9
BS213	λ papa lysogen of BS169	2 small cryptic plasmids	this study
E. coli			
JM101	Δ *lac pro, supE, thi,* F′ *traD36, proAB, lacI*^{q} ZΔ *M15*	none	26
pWR100	virulence plasmid from strain M90T		5
pHS4108	recombinant plasmid containing a 45 kb insert from pWR100		9
pACYC184	cloning vector Tc^r Cm^r		11

Materials and methods

Bacterial strains, plasmids and media. Bacterial strains and plasmids are listed in Table 2. Bacteria were routinely grown in L broth. For transfection and production of single stranded templates, strain JM101 was grown in SOB (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM $MgSO_4$ pH 6.8–7) and 2YT (1.6% tryptone, 1.6% yeast extract, 0.5% NaCl), respectively.

M13 cloning. Isolation of plasmid, purification and modification of DNA fragment, DNA ligation and transformation were carried out as described in Maniatis *et al.*[25] Plasmid pHS5753 was sonicated, treated with the Klenow fragment of DNA polymerase I (Genofit) in the presence of deoxyribonucleotides (Boehringer) for 16 h at 16°C, then fractionated by agarose gel electrophoresis. Fragments of 200–500 bp were electro-eluted and purified by chromatography on DEAE Sephacell (Pharmacia). DNA was ethanol-precipitated, retreated with the Klenow fragment of DNA polymerase I (Genofit) and T4 DNA polymerase (BRL), ligated to dephosphorylated *Sma*I-cleaved M13 mp8 RF DNA (Amersham) using T4 DNA ligase (Biolabs) for 16 h at 16°C and transfected into *Escherichia coli* strain JM101. Transfected bacteria were then spread on an agar medium in the presence of 5-bromo-4-chloro-3-indolyl-βD-galactopyranoside(X gal).

Screening of M13 recombinants. White M13 clones containing a DNA insert were replicated onto filters and screened by colony hybridization using plasmid vector pACYC184 labelled with ^{32}P as a probe. Phages that did not hybridize with this probe were selected for sequencing.

Sequencing technique and computer analysis. Preparation of single stranded DNA from individual plaques was performed as described by Messing.[26] The DNA sequence was determined by the dideoxy chain termination procedure[25] using 2′-deoxyadenosine 5′-[^{35}S][-thio] triphosphate (Amersham, 400 Ci/mmol) and buffer gradient gel.[28] Sequences were compiled and analysed using the programs of Staden[29–31] adapted by B. Caudron for the MV8000 computer of the Institut Pasteur Computer Center. Hydrophobicity profiles were calculated by the method of Kyte and Doolittle.[32]

Immunoblots. Whole bacterial extracts from BS169/pHS5753 were run on 0.1% sodium dodecyl sulfate (SDS)–12.5% polyacrylamide gels and blotted onto nitrocellulose filters as described by Burnette.[33] The protein loaded filters were treated as described by Fisher *et al.*[34] Diluted (1/250) convalescent serum from a monkey which had been orally infected with *S. flexneri* 2a was used to detect expression of immunogenic polypeptides.

Analysis of proteins expressed in minicells. Purification of minicells from 14 h L broth cultures was accomplished by differential centrifugation and three sucrose density gradient separations.[35] Purified minicells were labelled for 1 h with [^{35}S] methionine (50 μCi/ml, 800 Ci/mmol, Amersham). After washing, minicells were solubilized and extracts were run on 0.1% SDS–

12.5% acrylamide PAGE.[36] Fixation and fluorography of dried gels was performed as previously described.[37]

We wish to thank Thierry Garnier for his assistance in the utilization of the computer programs, and Armelle Phalipon and Catherine Gelin for their help. B.B. was supported by a fellowship from the "Ministère de la Recherche et de l'Enseignement Supérieur".

References

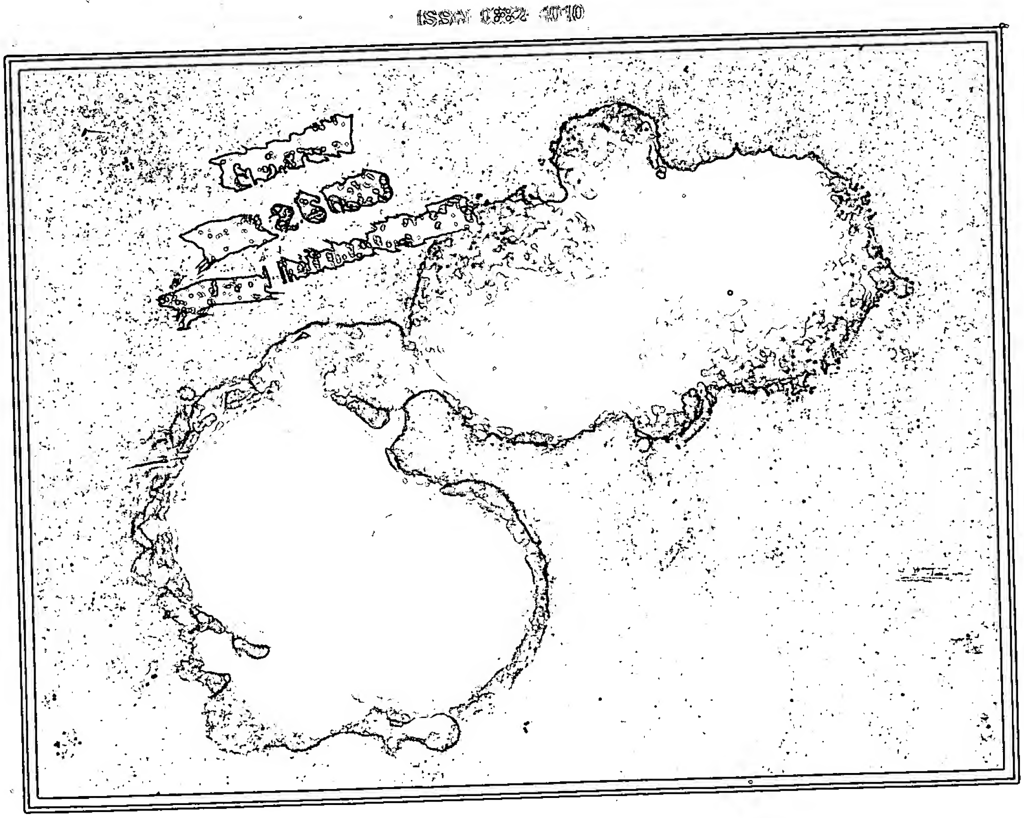
1. LaBrec EH, Schneider H, Magnani TJ, Formal SB. Epithelial cell penetration as an essential step in the patogenesis of bacillary dysentery. J Bacteriol 1964; 88: 1503–18.
2. Formal SB, LaBrec EH, Schneider H, Falkow S. Restoration of virulence to a strain of *Shigella flexneri* by mating with *Escherichia coli*. J Bacteriol 1965; 88: 835–8.
3. Formal SB, LaBrec EH, Kent TH, Falkow S. Abortive intestinal infection with an *Escherichia coli—Shigella flexneri* hybrid strain. J Bacteriol 1965; 89: 1374–82.
4. Sansonetti PJ, Kopecko DJ, Formal SB. *Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence. Infect Immun 1981; 34: 75–83.
5. Sansonetti PJ, Kopecko DJ, Formal SB. Involvement of a plasmid in the invasive ability of *Shigella flexneri*. Infect Immun 1982; 35: 852–60.
6. Sansonetti PJ, Hale TL, Dammin GJ, Kapfer C, Collins HH Jr, Formal SB. Alterations in the pathogenicity of *Escherichia coli* K12 after transfer of plasmid and chromosomal genes from *Shigella flexneri*. Infect Immun 1983; 39: 1392–402.
7. Hale TL, Oaks EV, Formal SB. Identification and antigenic characterization of virulence-associated, plasmid-coded proteins of *Shigella* spp. and enteroinvasive *Escherichia coli*. Infect Immun 1985; 50: 620–29.
8. Oaks EV, Hale TL, Formal SB. Serum immune response to *Shigella* protein antigens in rhesus monkeys and humans infected with *Shigella* spp. Infect Immun 1986; 53: 57–63.
9. Maurelli AT, Baudry B, d'Hauteville H, Hale TL, Sansonetti PJ. Cloning of plasmid sequences involved in invasion of HeLa cells by *Shigella flexneri*. Infect Immun 1985; 49: 164–71.
10. Baudry B, Maurelli AT, Clerc P, Sadoff JC, Sansonetti PJ. Localization of plasmid loci necessary for entry of *Shigella flexneri* into HeLa cells, and characterization of one locus encoding four immunogenic polypeptides. J Gen Microbiol 1987; 113: 3403–13.
11. Chang PAC, Cohen SN. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic plasmid. J Bacteriol 1978; 134: 1141–58.
12. Shine J, Dalgarno L. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc Natl Acad Sci USA 1974; 71: 1342–46.
13. Brenner DJ, Fanning GR, Miklos GV, Steigerwalt AG. Polynucleotide sequence relatedness among *Shigella* species. Int J Syst Bacteriol 1973; 23: 1–7.
14. Sansonetti PJ, Hale TL, Dammin GJ, Kapfer C, Collins HH Jr, Formal SB. Alterations in the pathogenicity of *Escherichia coli* K-12 after transfer of plasmid and chromosomal genes from *Shigella flexneri* Infect Immun 1983; 39: 1392–1402.
15. Rosenberg M, Court D. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu Rev Genet 1979; 13: 319–53.
16. Miller J, Reznikoff W. Eds. The operon. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. 1980.
17. Konigsberg W, Godson GN. Evidence for the use of rare codons in the *dna*G gene and the other regulatory genes of *Escherichia coli*. Proc Natl Acad Sci USA 1983; 80: 687–91.
18. Lockman H, Kaper JB. Nucleotide sequence analysis of the A2 and B subunits of *Vibrio cholerae* enterotoxin. J Biol Chem 1983; 258: 13722–6.
19. Mekalanos JJ, Swartz DJ, Pearson GDN, Harford N, Groyne F, deWilde M. Cholera toxin genes: Nucleotide sequence, deletion analysis and vaccine development. Nature 1983; 306: 551–7.
20. Jackson MP, Newland JW, Holmes RK, O'Brien AD. Nucleotide sequence analysis of the structural for Shiga-like toxin I encoded by a bacteriophage 933J from *Escherichia coli*. Microbiol Pathol 1987; 2: 147–53.
21. Calderwood SB, Auclair F, Donohue-Rolfe A, Keusch GT, Mekalanos JJ. Nucleotide sequence of the Shiga-like toxin genes of the *Escherichia coli*. Proc Natl Acad Sci USA 1987; 84: 4364–8.
22. Pugsley A, Schwartz M. Export and secretion of proteins by bacteria. FEMS Microbiol Rev 1985; 32: 3–38.
23. Sakai T, Sasakawa C, Makino S, Yoshikawa M. DNA sequence and product analysis of the *vir*F locus responsible for the Congo red binding and cell invasion in *Shigella flexneri* 2a. Infect Immun 1986; 54: 395–402.
24. Cantor CR, Schimmel PR. The behavior of biological macromolecules. In Biophysical Chemistry part III. Freeman WH and Co Ed. San Francisco. 1979.
25. Maniatis T, Fritsch EF, Sambrook J. Molecular Cloning: a Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. 1982.
26. Messing J. New M13 vectors for cloning. Methods Enzymol 1983; 101: 20–78.

27. Sanger F, Nicken S, Coulson AR. DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 1977; 74: 5463–7.
28. Biggin MD, Gibson TJ, Hong GF. Buffer gradient gels and 35S-label as an aid to rapid DNA sequence determination. Proc Natl Acad Sci USA 1983; 80: 3963–5.
29. Staden R. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res 1980; 6: 2601–10.
30. Staden R. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res 1980; 8: 3673–94.
31. Staden R. An interactive graphics program for comparing and aligning nucleic acid or amino acid sequences. Nucleic Acids Res 1982; 10: 141–56.
32. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157: 105–32.
33. Burnette WN. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radio-iodinated protein A. Anal Biochem 1981; 112: 195–203.
34. Fischer PA, Berrios M, Blobel G. Isolation and characterization of a proteinaceous subnuclear fraction composed of nuclear matrix, peripheral lamina, and nuclear pore complexes from the embryos of *Drosophila melanogaster*. J Cell Biol 1982; 92: 674–86.
35. Gemski P Jr, Griffin DE. Isolation and characterization of minicell-producing mutants of *Shigella* spp. Infect Immun 1980; 30: 104–11.
36. Laemli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4 phage. Nature 1970; 227: 680–5.
37. Hale TL, Sansonetti PJ, Schad PA, Austin S, Formal SB. Characterization of virulence plasmids and virulence-associated outer membrane proteins in *Shigella flexneri* 2a, *Shigella sonnei*, and *Escherichia coli*. Infect Immun 1983; 40: 340–50.

MICROBIAL PATHOGENESIS

Molecular and Cellular Biology of Infectious Disease

ISSN 0882-4010



ACADEMIC PRESS
(Harcourt Brace Jovanovich, Publishers)
London Orlando San Diego New York
Toronto Montreal Sydney Tokyo